

Analytical computation of frequency distributions of path-dependent processes by means of a non-multinomial maximum entropy approach

Rudolf Hanel¹, Bernat Corominas-Murtra¹, and Stefan Thurner^{1,2,3*}

¹*Section for Science of Complex Systems, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria*

²*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA and*

³*IIASA, Schlossplatz 1, 2361 Laxenburg, Austria*

(Dated: September 1, 2016)

Path-dependent stochastic processes are often non-ergodic and observables can no longer be computed within the ensemble picture. The resulting mathematical difficulties pose severe limits to the analytical understanding of path-dependent processes. Their statistics is typically non-multinomial in the sense that the multiplicities of the occurrence of states is not a multinomial factor. The maximum entropy principle is tightly related to multinomial processes, non-interacting systems, and to the ensemble picture; It loses its meaning for path-dependent processes. Here we show that an equivalent to the ensemble picture exists for path-dependent processes, such that the non-multinomial statistics of the underlying dynamical process, by construction, is captured correctly in a functional that plays the role of an entropy. We demonstrate this for self-reinforcing Pólya urn processes that explicitly break multinomial structure. We demonstrate the new method by computing frequency and rank distributions of Pólya urn processes. For the first time we are able to use detailed microscopic update rules of a path-dependent process to construct a non-multinomial entropy functional, that, when maximized, predicts the time-dependent distribution function.

Keywords: Pólya urns, statistical mechanics, maximum entropy principle, relative entropy, information divergence

I. INTRODUCTION

“It seems questionable whether the Boltzmann principle alone, meaning without a complete [...] mechanical description or some other complementary description of the process, can be given any meaning.” Einstein’s famous critical comment on the completeness of Boltzmann entropy, [1], is still thought provoking. For ergodic systems with a well defined number of states this critique has turned out to be of minor relevance. Here we show that Einstein’s observation becomes relevant again when dealing with non-ergodic and path-dependent systems. In fact we will demonstrate the possibility to directly construct “entropic functionals” from the microscopic properties determining the dynamics of a large class of non-ergodic processes. By construction, maximisation of the resulting functionals leads to correct predictions of statistical properties of non-ergodic processes.

For ergodic processes it is possible to replace time-averages of observables by their ensemble-averages, which leads to a tremendous simplification of computations. In particular, the use of entropy and its maximization under constraints has become a standard procedure for understanding and predicting distribution functions of large systems in equilibrium. In such applications the notion of *ergodicity* is crucial, meaning that observed probabilities essentially coincide with prior probabilities. This is certainly true for systems composed of independent particles or processes with independent increments, where the states of the independent components follow

a multinomial statistics. The multinomial statistics for a system with W states is captured by the Shannon entropy functional [2], $H(p) = -\sum_{i=1}^W p_i \log p_i$, which is nothing but the logarithm of the multinomial factor, $\sum_{i=1}^W p_i \log p_i \sim \frac{1}{N} \log \binom{N}{k}$, where N is the number of iterations, and $\binom{N}{k} = N! / \prod_{i=1}^W k_i!$ (e.g. compare [3]).

Maximization of entropy under constraints therefore is a way to find the most likely distribution function that one will observe when measuring a system, provided that it follows a multinomial statistics. Constraints represent knowledge about the system, such as moments of the distribution function. The set of parameters characterizing a system, including prior probabilities and constraint related parameters, we denote by θ . Suppose we perform an experiment and measure every state i , say k_i times. The total number of measurements is $N = \sum_i k_i$. We call $k = (k_1, \dots, k_W)$ the *histogram* of the experiment, $p = k/N = (p_1, \dots, p_W)$ is its *relative frequency* distribution. Denoting the probability to measure a specific histogram by $P(k|\theta, N)$, the most likely histogram \hat{k} , that maximizes $P(k|\theta, N)$ is the optimal predictor or the so-called *maximum configuration*. For a multinomial distribution function, $P(k|\theta, N) = \binom{N}{k} q_i^{k_i}$, where q_i are the prior probabilities (or biases), the functional that is maximized is $\psi(p|\theta) = H(p) + \sum_i p_i \log q_i$, which is called the *relative entropy* or Kullback-Leibler divergence [4]. The term $H(p)$ is Shannon entropy, the term that depends on q is called *cross-entropy* and is a linear functional in p . Re-parametrizing $q_i = \exp(-\beta \varepsilon_i)$, one gets the standard max-ent functional

$$\psi(p|\theta, N) = H(p) - \beta \sum_i p_i \varepsilon_i \quad \left[\sum_i p_i = 1 \right] \quad . \quad (1)$$

*Electronic address: stefan.thurner@meduniwien.ac.at

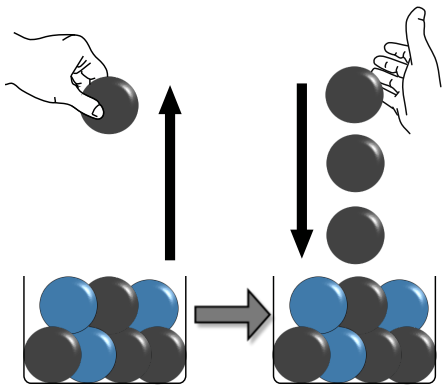


FIG. 1: Schematic illustration of a Pólya process. When a ball of a certain color is drawn, it is replaced by $1 + \delta$ balls of the same color. Then the next ball is drawn and the process is repeated for N iterations. Here $\delta = 2$. This reinforcement process creates a history-dependent dynamics. The configurations obtained after successive iterations have non-multinomial structure.

Maximization of this functional with respect to p yields the expected distribution functions; this is sometimes called the *maximum entropy principle* (MEP).

Clearly, systems composed of independent components follow a multinomial statistics. Note that a multinomial statistics for $P(k|\theta, N)$ is also a direct consequence of working with ensembles of statistically independent systems. In this case the multinomial statistics $P(k|\theta, N)$ reflects the ensemble property and is not necessarily a property of the system itself. Therefore $H(p)$ only has physical relevance for systems that consist of sufficiently independent elements. For path-dependent processes, where ensemble- and time-averages typically yield different results, $H(p)$ remains the entropy of the ensemble picture, but ceases to be the “physical” entropy that captures the time evolution of a path-dependent process. Obviously, using a multinomial entropy functional for path-dependent processes that are inherently non-multinomial (break multinomial symmetry) is nonsensical.

Surprisingly, the possibility that non-multinomial max-ent functionals can be constructed for path-dependent processes seems to have been overlooked. In [5] it was noticed that a very particular class of strongly correlated non-Markovian random walks, where the multiplicity of event sequences is no longer multinomial, can be associated with generalized entropy functionals, that emerge as a consequence of the violation of the composition axiom of Khinchin [6]. In this work we generalize this finding to multi-state *Pólya urn processes*, [7, 8], and thus open the way for general treatment of path-dependent, and non-equilibrium processes. We demonstrate how a max-ent approach can be successfully constructed for a wide class of non-multinomial processes.

In multi-state Pólya processes, once a ball of a given color is drawn from an urn, it is replaced by a number of δ balls with the same color. They are self-reinforcing, path-dependent processes that display the *the rich get*

richer or the *the winner takes all* phenomenon. Pólya urns are related to the beta-binomial distribution, Dirichlet processes, the Chinese restaurant problem, and models of population genetics. Their mathematical properties were studied in [9, 10], extensions and generalizations of the concept are found in [11, 12], applications to limit theorems in [13–15]. Pólya urns have been used in a wide range of practical applications including response-adaptive clinical trials [16], tissue growth models [17], institutional development [18], computer data structures [19], resistance to reform in EU politics [20], aging of alleles and Ewens’s sampling formula [21, 22], image segmentation and labeling [23], and the emergence of novelties in evolutionary scenarios [24, 25]. Another notion of Pólya-divergence was recently defined [26] in the context of Sanov’s theorem [27]. Our constructive approach proceeds along very different lines, yielding a functionally distinct notion of divergence, i.e. relative entropy.

II. NON-MULTINOMIAL MAX-ENT FUNCTIONALS

From a given class of processes X we select a particular process $X(\theta)$, specified by a set of parameters, θ . Running the processes $X(\theta)$ for N consecutive iterations produces a sequence of *observed states* $x(\theta, N) = [x_1, \dots, x_N]$, where each x_n takes a value from W possible states. As before we assume the existence of a most likely histogram \hat{k} , that maximises $P(k|\theta, N)$. To construct a max-ent functional for X , one has to conveniently rescale $P(k|\theta, N)$, which happens in two steps. First, we define $\Psi(p|\theta, N) \equiv \log P(Np|\theta, N)$. Second, if Ψ scales with N like the scaling function $\phi(N) = N^c$, for some constant c , we *identify* the max-ent functional with $\psi(p|\theta, N) \equiv \Psi(p|\theta, N)/\phi(N)$. For the necessity of such a scaling function, see [5]. Obviously, if \hat{k} maximises $P(k|\theta, N)$ then $\hat{p} = \hat{k}/N$ maximises $\psi(p|\theta, N)$, with $\sum_i p_i = 1$.

III. MAX-ENT FUNCTIONAL FOR PÓLYA URNS

In urn models states i are represented by the colors balls can have. The likelihood of drawing a ball of color i is determined by the number of balls contained in the urn. Initially the urn contains a_i balls of color $i = 1, \dots, W$. The *initial prior probability* to draw a ball of color i is given by $q_i = a_i/A_0$, where $A_0 = \sum_i a_i$ is the total number of balls initially in the urn. Balls are drawn sequentially from the urn. Whenever a ball of color i is drawn, it is put back into the urn and another δ balls of the same color are added. This defines the multi-state Pólya process [7]. A particular Pólya process is fully characterised by the parameters, $\theta = (q_1, \dots, q_W; A_0, \delta)$. Drawing without replacement is the *hypergeometric* process,

drawing with replacement ($\delta = 0$), is the *multinomial* process.

If $\delta > 0$, after N trials there are $a_i(N) = a_i + \delta k_i$ balls of color i in the urn ($a_i = a_i(0)$). The total number of balls is $A(N) = \sum a_i(N) = A_0 + N\delta$, and the probability to draw a ball of color i in the $(N+1)$ th step is

$$p(i|k, \theta) = \frac{a_i(N)}{A(N)} = \frac{a_i + k_i \delta}{A_0 + N\delta}, \quad (2)$$

which depends on the history of the process in terms of the histogram k . With $x(0) = []$ the *empty sequence*, the probability of sampling sequence x can be computed

$$p(x|\theta) = \prod_{n=1}^N p(x_n|k(x(n-1)), \theta) = \frac{\prod_{i=1}^W a_i^{(\delta, k_i)}}{A_0^{(\delta, N)}}, \quad (3)$$

where the function $m^{(\delta, r)}$ is defined as

$$m^{(\delta, r)} \equiv m(m + \delta)(m + 2\delta) \cdots (m + (r-1)\delta). \quad (4)$$

Note that $m^{(\delta, r)}$ generalises the multinomial law,

$$\left(\sum_i a_i \right)^{(\delta, N)} = \sum_{\{k|N=\sum_i k_i\}} \binom{N}{k} \prod_{i=1}^W a_i^{(\delta, k_i)}, \quad (5)$$

and forms a one-parameter generalisation of powers m^r . For $\delta = 0$, $m^{(0, r)} = m^r$ and for $\delta = 1$, $m^{(1, r)} = (m + r - 1)!/(m - 1)!$.

The probability of observing a particular histogram k after N trials becomes

$$P(k|\theta, N) = \binom{N}{k} \frac{\prod_{i=1}^W a_i^{(\delta, k_i)}}{A_0^{(\delta, N)}}, \quad (6)$$

with $\sum_{\{k_i \geq 0 | \sum_i k_i = N\}} P(k|\theta, N) = 1$. Note that $P(k|\theta, N)$ is almost of multinomial form, it is a multinomial factor times a term depending on θ . One might conclude that the max-ent functional for Pólya processes is Shannon entropy in combination with a generalised cross-entropy term that depends on θ . However, this turns out to be wrong, since contributions from the generalised powers $m^{(\delta, r)}$ in equation (6) cancel the multinomial factor almost completely. To see this we first rewrite

$$\begin{aligned} a_i^{(\delta, k_i)} &= a_i(a_i + \delta) \cdots (a_i + (k_i - 1)\delta) \\ &= (a_i + \delta) \cdots (a_i + k_i \delta) \frac{a_i}{a_i + k_i \delta} \\ &= k_i! \delta^{k_i} \left(1 + \frac{a_i}{\delta}\right) \cdots \left(1 + \frac{a_i}{k_i \delta}\right) \frac{a_i}{a_i + k_i \delta} \\ &\sim k_i! \delta^{k_i} (k_i + 1)^{\frac{a_i}{\delta}} \frac{a_i}{a_i + k_i \delta}, \end{aligned} \quad (7)$$

where we use $\sum_{r=1}^s \frac{1}{r} \sim \log(s+1)$ and $1+y \sim \exp(y)$, which is valid for sufficiently small $y = a_i/\delta$, i.e. for sufficiently large δ . With the notation $\gamma \equiv \delta/A_0$ we obtain

$$P(k|\theta, N) = \frac{N^{-W-1}}{(\gamma + \frac{1}{N}) (1 + \frac{1}{N})^{\frac{1}{\gamma}}} \prod_{i=1}^W \frac{(p_i + \frac{1}{N})^{\frac{q_i}{\gamma}}}{\frac{p_i}{q_i} \gamma + \frac{1}{N}}. \quad (8)$$

Following the construction discussed above, we identify $\Psi(p|\theta, N) = \log P(pN|\theta, N)$ ($k = Np$), which no longer scales explicitly with N , but $\phi(N) = 1$ ($c = 0$), so that $\psi = \Psi$. We note that in fact all terms depend on the weights q_i . However, in the limit $N \rightarrow \infty$ we obtain the asymptotic result for the generalised "Pólya information divergence" for large N ,

$$\psi(p|\theta) = - \sum_{i=1}^W \log p_i + \frac{1}{\gamma} \sum_{i=1}^W q_i \log p_i. \quad (9)$$

Now one can identify those terms shown in equation (9) that do not depend on q and identify them as the "entropy" of the system. Thus the asymptotic Pólya "entropy" up to an additive constant, is given by,

$$H^{\text{Pólya}}(p) = - \sum_{i=1}^W \log p_i. \quad (10)$$

We observe that one cannot derive $H^{\text{Pólya}}(p)$ from the multiplicities of the system, which gets canceled by counter terms, as we have seen above. In addition, we note that the q dependent terms, $\sum_i q_i \log p_i$, in equation (9) play the role of the Pólya "cross-entropy", which is no longer linear in p .

maximising $\psi(p|\theta)$ with respect to p on $\sum p_i = 1$, either leads to the solution

$$p_i = \frac{1}{\zeta} (q_i - \gamma), \quad (11)$$

for $0 < p_i < 1$, or, if this can not be satisfied, to boundary solutions $p_i = 0$. ζ is a normalisation constant. There exist three scenarios:

- (A) For $\gamma < \min(q)$, equation (11) is the correct max-ent solution for all i (no boundary-solutions). The limit $\gamma \rightarrow 0$ provides the correct multinomial limit $p_i \rightarrow q_i$.
- (B) If $\max(q) > \gamma > \min(q)$, ψ gets maximal for those i with $q_i > \gamma$ and follows solution equation (11); those i where $q_i < \gamma$ are boundary-solutions, $p_i = 0$.
- (C) For $\gamma > \max(q)$ all p_i are boundary-solutions, meaning that one winner i takes it all, $p_i = 1$, while all other states have vanishing probability.

Since $\partial^2 \psi_{\text{Pólya}} / \partial p_i^2 < 0$ if $\gamma < q_i$, $\psi_{\text{Pólya}}$ is maximal and case (A) applies. If $q_i < \gamma$, Eq (11) becomes negative but also unstable and is replaced by a boundary solution: cases (B) and (C). The Pólya max-ent not only allows us to predict p_i from the initial prior probabilities q_i , it also identifies γ as the crucial parameter that distinguishes between the three regimes of Pólya urn dynamics. For sufficiently large but finite N , the analysis above is more involved but solvable. Up to an additive constant the *finite size* Pólya "entropy" can be identified with the term:

$$H^{\text{Pólya}}(p) = - \sum_{i=1}^W \log \left(p_i + \frac{1}{NW\gamma} \right). \quad (12)$$

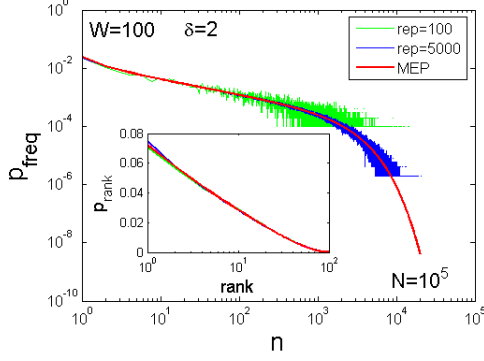


FIG. 2: Frequency distribution of a Pólya urn process and uniform initial conditions (red line), for $W = 100$, $\delta = 2$, $a_i = 1$ for all $i = 1$, and $N = 10^5$ steps. Simulations are shown for 100 (green) and 5000 (blue) repetitions of the process. Inset: Rank distributions of the max-ent result and the numerical realisations in semi-log scale.

In contrast to the asymptotic Pólya entropy, equation (10), the finite size Pólya entropy equation (12), yields well defined entropy values for normalised histograms p , even when some i will occur as boundary solutions with vanishing probability $p_i = 0$. The finite size cross-entropy then reads

$$H_{\text{cross}}^{\text{Pólya}}(p|q) = - \sum_i \left[\frac{1}{\gamma} q_i \log(p_i + 1/N) - \log \left(1 - \frac{1 - W q_i}{1 + \gamma W N p_i} \right) \right]. \quad (13)$$

Assuming uniformly distributed priors, $q_i = 1/W$ for all i , the max-ent result equation (11) correctly predicts uniformly distributed $\hat{p}_i = 1/W$, while observed distributions p may strongly deviate from this prediction. This result reflects the fact that despite the Pólya urn process being inherently unstable (e.g. winner takes all) with little chance of predicting who in particular will win, i.e. which color of balls will dominate the others, repeating the experiment many times every color of balls has the same chance to win (or biased according to the priors q). This discrepancy between ensemble average and time average makes it impossible to predict who in particular will win or loose in the course of time. However, using detailed information about the process one can predict how strongly winners wins. In particular one can construct a maximum entropy functional for predicting the time dependent frequency distribution of a process, i.e. the number of times one observes states i for n times. As a consequence one also can derive the rank distributions of the process, i.e. the frequency of observing balls of some color after ranking those frequencies according to their magnitude.

A. Rank and frequency distributions of Pólya urns

With the presented max-ent approach we now compute frequency distribution functions. Given the histogram $k = (k_1, k_2, \dots, k_W)$ is obtained after N iterations of the process, we define new variables,

$$n_z(k) = \sum_{i=1}^W \chi(k_i = z), \quad (14)$$

where χ is the characteristic function that returns 1 if the argument is true and 0 if false. $n_z(k)$ is the number of colors i that occur z times after running the Pólya process for N iterations. n_z is subject to the two constraints,

$$W = \sum_{z=0}^N n_z(k) \quad \text{and} \quad N = \sum_{z=0}^N n_z(k) z. \quad (15)$$

The probability of observing some $n = (n_1, \dots, n_N)$ is

$$\tilde{P}(n|\theta, N) = \sum_{\{k_i \geq 0 | n=n(k)\}} P(k|\theta, N). \quad (16)$$

Defining the relative frequencies $\pi_z = n_z/W$ and $\bar{p}_z = z/N$ we can construct the max-ent functional from $\tilde{P}(n|\theta, N)$. We identify $\tilde{\psi}(\pi|\theta, N) \equiv \log(\tilde{P}(n|\theta, N))/W$.

For the multinomial $P(k|\theta, N) = \binom{N}{k} \prod_i q_i^{k_i}$, and uniform priors $q_i = 1/W$ we find up to an additive constant,

$$\begin{aligned} \tilde{\psi}(\pi|\theta, N) = & - \sum_{z=0}^N \pi_z \log \pi_z \\ & - N \sum_{z=0}^N \pi_z \bar{p}_z \log \bar{p}_z. \end{aligned} \quad (17)$$

$\tilde{\psi}(\pi|\theta, N)$ has to be maximised subject to equation (15),

$$\sum_{z=0}^N \pi_z = 1 \quad \text{and} \quad \sum_{z=0}^N \pi_z \bar{p}_z = \frac{1}{W}, \quad (18)$$

so that we get the asymptotic solution for large W and large N , ($N \gg W \gg 1$),

$$\pi_z = \frac{\phi^z}{\zeta z!}. \quad (19)$$

This is the Poisson distribution, exactly as expected for multinomial processes. $\phi = N e^{-(1+\frac{\beta}{N})}$, ζ is a normalisation constant, and π_z gets maximal at $\hat{z} = \phi \sim N/W$.

For the Pólya urn with uniform priors we get from equation (16)

$$\tilde{P}(n|\theta, N) = \frac{1}{Z(\theta, N)} \frac{W!}{\prod_{z=0}^N n_z!} \prod_{z=0}^N \left[\frac{(\bar{p}_z + \frac{1}{N})^{\frac{1}{W\gamma}}}{\bar{p}_z + \frac{1}{\gamma W N}} \right]^{n_z}. \quad (20)$$

$Z(\theta, N)$ is the normalisation. Up to a constant the max-ent functional $\tilde{\psi}_{\text{Pólya}}(\pi|\theta, N) \equiv \log(P(n|\theta, N))/W$ is

$$\begin{aligned} \tilde{\psi}_{\text{Pólya}}(\pi|\theta, N) = & -\sum_{z=0}^N \pi_z \log(\pi_z) \\ & -\sum_{z=0}^N \pi_z \log\left(\bar{p}_z + \frac{1}{\gamma W N}\right) \\ & +\frac{1}{W\gamma} \sum_{z=0}^N \pi_z \log\left(\bar{p}_z + \frac{1}{N}\right). \end{aligned} \quad (21)$$

maximising $\tilde{\psi}_{\text{Pólya}}(\pi|\theta, N)$ under the conditions of equation (18) provides the frequency distribution of the Pólya process for uniform priors,

$$\pi_z = \frac{1}{\zeta} \phi^z \frac{(z+1)^{\frac{1}{W\gamma}}}{z + \frac{1}{\gamma W}}, \quad (22)$$

with $\phi = \exp(-\beta)$, and normalisation $\zeta = \exp(1 + \alpha)N^{\frac{1}{W\gamma}-1}$.

The rank distribution of states, $f(r)$, can now be obtained as follows. $r = 1$ is the state that occurs most frequently, $r = W$ is the least occupied state. For $r = 1, \dots, W$ we define intervals $[t_{r+1}, t_r]$ with $t_1 = N$ and $t_{W+1} = 0$, such that $\sum_{t_{r+1} \leq z < t_r} \pi_z \sim 1/W$. To find t_i we substitute sums by integrals and get

$$\frac{1}{W} = \int_{t_{r+1}}^{t_r} dz \pi_z \quad \text{and} \quad f(r) = \frac{W}{N} \int_{t_{r+1}}^{t_r} dz \pi_z z. \quad (23)$$

Results for the frequency distributions for $a_i = 1$, $W = 100$, and $\delta = 2$ are shown in Fig. (2), together with a numerical simulation for the same process. The inset shows the rank distribution. The Pólya max-ent predicts frequency and rank distribution extremely well.

The above results were derived under the assumption that $\gamma > 0$ is sufficiently large. By numerical simulation we find that the solution equation (22) also works remarkably well for very small values of γ , if the value of γ in equation (22) is appropriately renormalised, $\gamma \rightarrow \gamma_0$. In particular for $\gamma = 0$ (multinomial process) we sample the Poisson distribution function, equation (19). The Pólya max-ent solution recovers the Poisson distribution extremely well if $\gamma = 0 \rightarrow \gamma_0(W, N) = 1/(N + 3W)$. In this sense the Pólya max-ent remains adequate in the limit of small γ .

IV. DISCUSSION

Pólya urns offer a transparent way to study self-reinforcing systems with explicit path-dependence.

Based on the microscopic rules of the process, we constructively derive the generalised information divergence ψ which acts as the corresponding non-multinomial max-ent functional. This provides us with an alternative to the *ensemble approach* for path-dependent processes that is able to predict the statistics of the system. The maximisation of the functional leads to an equivalent to the classical *maximum configuration* approach, which by definition predicts the most likely distribution function. In this sense maximum configuration predictions are optimal, and can be used to understand even details of the statistics of path-dependent processes, such as their frequency and rank distributions.

It is interesting to note that the functional playing the role of the entropy in the Pólya processes violates at least two of the four classic information theoretic (Shannon-Khinchin) axioms which determine Shannon entropy [6]. Even more, for the finite size Pólya entropy, three of the four axioms are violated. This indicates that the classes of generalised entropy functionals that are useful for a max-ent approach may be even larger than expected [28, 29]. One might speculate that in this sense the classic information theoretic axioms are too rigorous, when it comes to characterising information flow and phase space structure in non-stationary, path-dependent, processes. The observation that each particular class of non-multinomial processes requires a matching max-ent functional that can in principle be constructed from the generative rules of a process, opens the applicability of max-ent approaches for a wide range of complex systems in a meaningful way. The generalized max-ent approach in this sense responds to Einsteins critique of the Boltzmann entropy in a natural way.

Finally we note the implications for statistical inference with data from non-multinomial sources, which implicitly involves the estimation of the parameters θ that determine the process that generates the data. In a max-ent approach this is done by fitting classes of curves to the data, that are consistent with the max-ent approach. For doing this, the nature of the process needs to be known. For path-dependent processes, which are non-multinomial by nature, entropy will no longer be Shannon entropy H , and the information divergence will no longer be the Kullback-Leibler divergence.

[1] Einstein A 1910 Theorie der Opaleszenz von homogenen Flüssigkeiten und Flüssigkeitsgemischen in der Nähe des kritischen Zustandes. *Ann. d. Phys.* **33**, 1275.

[2] Shannon C E 1948, A Mathematical Theory of Communication, *Bell Syst. Tech. J.* **27** 379-423, 623-656.

[3] Jaynes E T 1968, Prior Probabilities, *IEEE Trans Sys*

- Sci and Cybernetics* **4** 227-241.
- [4] Kullback S, and Leibler R A 1951, On information and sufficiency, *Ann. Math. Stat.* **22** 79-86.
 - [5] Hanel R, Thurner S, and Gell-Mann M 2014, How multiplicity of random processes determines entropy: derivation of the maximum entropy principle for complex systems, *Proc. Nat. Acad. Sci. USA* **111** 6905-6910.
 - [6] Khinchin A I 1957, *Mathematical foundations of information theory*, (Dover Publ., New York).
 - [7] Eggenberger F, G. Pólya G 1923, Über die Statistik verketteter Vorgänge, *Z. Angew. Math. Mech.* **1** 279-289.
 - [8] Pólya G 1930, Sur quelques points de la théorie des probabilités, *Ann. Inst. Henri Poincaré* **1** 117-161.
 - [9] Wallstrom T C 2012, The equalization probability of Pólya urn, *Am. Math. Mon.* **119** 516-518.
 - [10] Johnson N L and Kotz S 1977, Urn Models and Their Application: An Approach to Modern Discrete Probability Theory, In *Urn models and their application. An approach to modern discrete probability theory*. (John Wiley, New York).
 - [11] Mahmoud H 2008, Pólya Urn Models, *Texts in Statistical Science*, (Chapman & Hall/CRC Texts in Statistical Science, Taylor and Francis Ltd, Hoboken, NJ).
 - [12] Kotz S, Mahmoud H, and Robert P 2000, On generalised Pólya urn models, *Stat. Prob. Lett.* **49** 163-173.
 - [13] Janson S 2004, Functional limit theorems for multitype branching processes and generalised Pólya urns, *Stoch. Proc. Appl.* **110** 177-245.
 - [14] Smythe R T 1996, Central limit theorems for urn models, *Stoch. Proc. Appl.* **65** 115-137.
 - [15] Gouet R 1993, Martingale Functional Central Limit Theorems for a generalised Pólya Urn, *Ann. Prob.* **21** 1624-1639.
 - [16] Tolusso D and Wang X 2011, Interval estimation for response adaptive clinical trials, *Comput. Stat. Data Anal.* **55** 725-730.
 - [17] Binder B J and Landman K A 2009, Tissue growth and the Pólya distribution, *Aust. J. Eng. Edu.* **15** 35-42.
 - [18] Crouch C and Farrell H 2004, Breaking the Path of Institutional Development? Alternatives to the New Determinism, *Ratio. Soc.* **16** 5-43.
 - [19] Bagchi A and Pal A K 1985, Asymptotic Normality in the generalised PólyaEggenberger Urn Model, with an Application to Computer Data Structures, *SIAM J. Algeb. Disc. Meth.* **6** 394-405.
 - [20] Geppert T 2012, EU-Agrar- und Regionalpolitik, Wie vergangene Entscheidungen zukünftige Entwicklungen beeinflussen - Pfadabhängigkeit und die Reformfähigkeit von Politikfeldern, *PhD Thesis*, (University of Bamberg Press, Bamberg).
 - [21] Donnelly P 1986, Partition structures, Pólya urns, the Ewens sampling formula, and the ages of alleles, *Theor. Popul. Biol.* **30** 271-288.
 - [22] Hoppe F M 1984, Pólya-like urns and the Ewens' sampling formula, *J. Math. Biol.* **20** 91-94.
 - [23] Banerjee A, Burlina P, and Alajaji F 1999, Image segmentation and labeling using the Pólya urn model, *IEEE Trans. Image. Proc.* **8** 1243-1253.
 - [24] Alexander J M, Skyrms B, and Zabell S 2012, Inventing new signals, *Dyn. Games Appl.* **2** 129-145.
 - [25] Tria F, Loreto V, Servedio V D P, and Strogatz S H 2014, The dynamics of correlated novelties, *Sci. Rep.* **4**:5890.
 - [26] Grendar M, and Niven R K 2010, The Pólya information divergence, *Info. Sci.* **180** 4189-4194.
 - [27] Sanov I N 1957, On the probability of large deviations of random variables, *Mat. Sbornik* **42** 11-44.
 - [28] Hanel R and Thurner S 2011, A comprehensive classification of complex statistical systems and an ab initio derivation of their entropy and distribution function, *EPL* **93**:20006.
 - [29] Hanel R and Thurner S 2011, When do generalised entropies apply? How phase space volume determines entropy, *EPL* **96**:50003.